

## 超早期診断のための呼気データ解析

情報科学科 玉置 翔

指導教員：作村 諭一

## 1 はじめに

現在、日本人の「ガン」による死亡者が年々増加してきている。厚生労働省の統計では、1981 年ごろから心疾患や、脳血管疾患などの症状を引き離し、死亡率は上昇の一步をたどっている。厚生労働省人口動態統計によると、平成 19 年の年間死亡者数は 110 万 8,334 人である。このうち、33 万 6,468 人 (男性:20 万 2,743 人、女性:13 万 3,725 人) が、ガンにより死亡している。このため、「日本人の 3 人に 1 人がガンで死亡している。」と言える。また、ガンによる死亡原因のトップとして肺ガンが増加している。

ガンの死亡率を低下させるための試みとして、ガンの予防、早期発見、治療法の確立があげられる。ガンの早期発見のネックとなるのは、内視鏡検査などで肉体的負担が大きいことである。また、従来のレントゲン検査においては、放射能による肉体的負担だけでなく、まだ小さいガンは発見されにくいという特徴もある [1]。そのため、呼気という簡易な手段での、ガンの早期発見法の確立は意義のあるものであると言える。

本研究では、ベイズの定理を用いて呼気データを解析し、患者のガン判定を行うことを目的とする。

## 2 ベイズの定理

ベイズの定理は、パターン認識や機械学習においてよく用いられており、近年の重要な手法として多用されている。ベイズの定理は以下で定義される。

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (1)$$

左辺の  $P(B|A)$  は事後確率であり、A を観測した後に B に関する不確実性を評価することを可能とする。また、右辺の  $P(A|B)$  はデータ集合 A に対する評価であり、A の尤度関数である。これは、A のパラメータを固定した際に観測されたデータ集合がどれくらい起こりやすいかを表している。

尤度の定義から、ベイズの定理は言葉で書けば

$$(\text{事後確率}) \propto (\text{尤度}) \times (\text{事前確率}) \quad (2)$$

となる。

また、ベイズの定理 (1) 式の分母は、左辺の事後分布が厳密に確率密度となっており、積分すると 1 になることを保障する規格化定数である [2]。

## 3 実験手法

ベイズ推定を呼気データに応用する。(2) 式に当てはめて考えていく。事前確率として与えられているのは、ガン患者、健常者である確率である。今回の研究では、どちらも同じ確からしさで呼気データを採取したものとし、解析を行う。

尤度については、ガン患者と健常者のガスを固有のしきい値によって二値化した結果のガス検出確率の積として評価する。ガスは複数あるが、採取した人の病状が決まれば各成分の検出確率は独立であるため、掛け合わせることが可能である。

以上の事前確率と尤度から、ガンである確率を計算し、ガン判定を行う。

## 4 結果考察

ガン判定をあげるために、ガスに対しガン確率の計算に大きな影響をもつ順に順位付けを行った。その中から上位のガスをいくつか使うことによって判定率の変化を示すものが図 1、図 2 である。なお、ガン患者、健常者合わせて 111 名に対して、ガン判定を行った。判定において重要ではないガスを抜いていることにより、判定率が極めて高いと言える。また、複数のガスを併用して求めた結果の方が優秀な値を出していることが分かる。

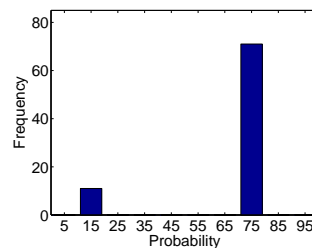


図 1 ガス 1 個を使った結果

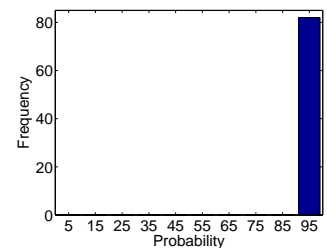


図 2 ガス 8 個を使った結果

また、図 3 では、しきい値を全ガスで同一の値にするのではなく、ガス毎に変更し、適する値にした場合を示す。しきい値設定は、まずガスの検出量のヒストグラムを健常者、ガン患者で中央値をとった。その中央値の差を 10 分割し、ガス 1 つで判定した結果がよかったものを使用した。ガン判定は全体として 7 割ほどの正答率となった。患者によって 100 % 近い判定率もあるが、完全に誤判定してしまっている患者もいることが分かる。これは、通常の人より呼気の量が極端に多いまたは、少ない患者がいることが原因であると推測される。

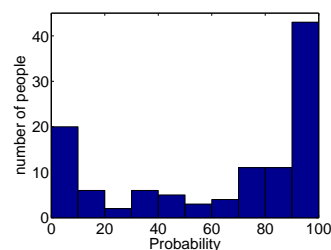


図 3 ガン判定結果

## 5 おわりに

呼気から、ガンの判定がある程度可能であることを示した。現在の判定率では実際に使うことができるレベルにはいたっていない。ただし、データが増加すればしきい値が安定し正答率も増加すると考えられる。今後は、患者によって呼気全体の検出量に差の問題や、曖昧な患者を判定する方法を考える必要がある。

## 参考文献

- [1] <http://kotobank.jp/word/がんの死亡率>  
 [2] C.M. ピショップ『パターン認識と機械学習 上』丸善出版,2012